Noname manuscript No.
(will be inserted by the editor)

Short hyperuniform random walks

Emanuele Casini · Gérard Le Caër · Andrea Martinelli

Received: date / Accepted: date

Abstract Random walks of two steps, with fixed sums of lengths of 1, taken into uniformly random directions in d-dimensional Euclidean spaces ($d \geq 2$) are investigated to construct continuous step-length distributions which make them hyperuniform. The endpoint positions of hyperuniform walks are spread out in the unit ball as the projections in the walk space of points uniformly distributed on the surface of the unit hypersphere of some k-dimensional Euclidean space ($k > d$). Unique symmetric continuous step-length distributions exist for given d and k, provided that $d < k < 2d$. The walk becomes uniform on the unit ball when $k = d + 2$. The symmetric densities reduce to

Emanuele Casini, Dipartimento di Scienze ed Alta Tecnologia, Universitá degli Studi dell'Insubria, Via Vallegio 5, 22100 Como, Italy. E-mail: emanuele.casini@uninsubria.it

Gérard Le Caër, Institut de Physique de Rennes, UMR UR1-CNRS 6251, Université de Rennes I, Campus de Beaulieu, Bâtiment 11A, 35042 Rennes Cedex, France. E-mail: gerard.le-caer@univ-rennes1.fr

Andrea Martinelli, Dipartimento di Scienze ed Alta Tecnologia, Universitá degli Studi dell'Insubria, Via Vallegio 5, 22100 Como, Italy. E-mail: andrea.martinelli@uninsubria.it

simple polynomials for uniform random walks and are mixtures of two pairs of asymmetric beta distributions.

Keywords Dirichlet random walks · Random flights · beta distribution · Hyperuniform distribution

PACS 05.40.Fb · 02.50.Cw

1 Introduction

A planar random walk (RW) made of a sequence of n steps of unit lengths taken into uniformly random directions was defined by Pearson in 1905 [30,31]. In spite of its idealized feature, the Pearson's RW finds various applications and was, for instance, used recently to characterize the cosmic microwave background [35,16]. Generalizations of Pearson's random walk involve space dimensions higher than two, changes of step length distributions, deviations of step orientations from a uniform repartition and introduction of correlations between steps. These extended Pearson's random walks are useful in diverse fields such as physics, biology, ecology (see [4,5,8,34,36,38] and references therein).

1.1 Dirichlet random walks

A large family of random walks of n steps in d–dimensional Euclidean spaces ($d \geq 2$) maintains the independence between successive steps and the uniformity of their orientations but step lengths are allowed to vary according to some continuous probability law. Recent studies include an additional modification in the form of a constraint on the sum of step lengths [4,8,10, 14,15,21–24]. These constrained step lengths are the components of a random vector $\boldsymbol{L}_{(n)} = (L_1, L_2, \cdots, L_n)$ and their sum is fixed to a value S, $\sum_{i=1}^{n} L_i = S = constant$. In almost all cases considered hereafter, S will be taken as equal to 1 without loss of generality. This $\boldsymbol{L}_{(n)}$ belongs to the

$(n-1)$ unit simplex. The previous constrained walks are related, in some of their aspects, to random flights performed in d–dimensional Euclidean spaces by particles which fly at a constant and finite speed c in some direction until they choose instantaneously, at random times, a new direction according to some probability law and fly again at speed c (see for istance [10–12, 18–20, 29, 32–34, 38]). At a given time t, the lengths of all flights are identical and equal to ct. Therefore, the conditional probability density function (pdf) of the position of a particle which flies during t, given that it underwent $n-1$ changes of direction, is identical, possibly after rescaling, to the pdf of the end-point position of constrained walks of n steps. The problem of the step length distribution of the considered family is related to the broken stick problem, i.e. the problem of the random splitting of a unit interval.

Almost all previous studies chose the Dirichlet distribution as the distribution of the random vector $\boldsymbol{L}_{(n)}$. The multivariate Dirichlet distribution, which is applied for instance to model fragmentation or compositional data [2], is conveniently defined from gamma distributed random variables (see Appendix A). The Dirichlet distribution of $\boldsymbol{L}_{(n)}$ is,

$$f_L\left(l_1, l_2, \ldots, l_{n-1}\right) = \frac{\Gamma\left(\sum_{i=1}^{n} q_i\right)}{\prod_{i=1}^{n} \Gamma\left(q_i\right)} \left(1 - \sum_{i=1}^{n-1} l_i\right)^{q_n} \prod_{i=1}^{n-1} l_i^{q_i - 1},$$

where $l_i > 0, i = 1, \ldots, n-1$, $\sum_{i=1}^{n-1} l_i \leq 1$. This distribution, denoted here as $D(q_1, q_2, .., q_n)$, depends on a n–dimensional vector of positive parameters $q(n) = (q_1, q_2, .., q_n)$ which may be looked at as being the shape parameters of the associated gamma distributions (see Appendix A), a name which may be kept hereafter. When all shape parameters are equal to $q > 0$, the

Dirichlet pdf becomes symmetric i.e. invariant under permutations. The related n–step random walks in d–dimensional Euclidean spaces [10, 14, 15, 21–24] are named "Dirichlet random walks" and will be denoted hereafter either as $W(d, n, \boldsymbol{q}_{(n)}) \equiv W\left(d, n, (q_1, q_2, \cdots, q_n)\right)$ or simply $W(d, n, q)$ when all Dirichlet parameters are equal to $q > 0$. The early studies of Dirichlet random walks dealt with a symmetric distribution for which $q = 1$. This case arises for instance when particles move in random environments and undergo elastic collisions at uniformly distributed point obstacles [10]. The associated random flights or walks have exponential distributions of step lengths with equal scale parameters. In addition, the sum of step lengths was constrained to be equal to 1 [10]. The initial impetus for imposing this condition was to find couples (d, n) for which the endpoints of such n-step walks, $W(d, n, 1)$, are uniformly distributed on the d–dimensional unit ball B_1^d [10]. Throughout the paper, B_r^d will designate the d-dimensional Euclidean ball of radius r centered at the origin.

1.2 Definition of hyperuniform random walks

The projections on the walk space $\mathbb{R}^d$ of points uniformly distributed on the surface of a unit sphere of a k-dimensional Euclidean space, whose dimension k is larger than d, have a radial density which is given by [9, 26]:

$$p_{d,k}(r) = \frac{2\Gamma(k/2)}{\Gamma(d/2)\Gamma((k-d)/2)} r^{d-1}(1-r^2)^{\frac{k-d-2}{2}}, \, r \in [0, 1]. \tag{1}$$

The latter projections are uniformly distributed on the unit ball B_1^d if and only if $k = d + 2$ as then $p_{d,d+2}(r) = dr^{d-1}$. The dimension k will be named from now on the "hyperspace dimension".

As proposed by Letac and Piccioni [24], this property, which was named "hypersherical uniform" by Le Caër [21], will be abbreviated hereafter to "hyperuniform". A n-step random walk in $\mathbb{R}^d$ $(n \geq 2)$, denoted henceforth $HU_d(k)$, is said to be hyperuniform of type $k > d$ if the endpoint of the walk is scattered as is the projection of a point uniformly distributed on the surface of the unit hypersphere of $\mathbb{R}^k$. Then, the pdf of the distance between the origin and the endpoint of a hyperuniform random walk $HU_d(k)$ is by definition given by equation (1). The endpoint of the hyperuniform random walk $HU_d(d+2)$ is then uniformly distributed on the ball B_1^d.

1.3 Hyperuniform Dirichlet random walks

Two families of Dirichlet random walks in $\mathbb{R}^d$ with symmetric step length distributions $D(q^{(i)}, q^{(i)}, \ldots, q^{(i)}), i = 1, 2$, and only two, exhibit the hyperuniform property for any number of steps $n \geq 2$ (a one–step random walk is trivially hyperuniform with $k = d$) [21]. The shape parameter $q^{(i)}$ depends solely on the walk space dimension d. By contrast, the hyperspace dimension $k^{(i)}$ depends both on d and on n. These parameters are [21]:

$$\begin{cases} (F_1) : q^{(1)} = d - 1, k^{(1)} = n(d-1) + 1 \quad (d \geq 2) \\ (F_2) : q^{(2)} = d/2 - 1, k^{(2)} = n(d-2) + 2 \ (d \geq 3) \end{cases} \tag{2}$$

The focus of the present paper will be on two–step random walks (Section 1.4). For two–step Dirichlet random walks, the values of $k^{(1)}$ and of $k^{(2)}$ are respectively $2d-1$ and $2d-2$.

We notice in passing that the pdf of the distance from the origin to the endpoint is simply derived for the two–step Dirichlet random walks $W(d,2,q)$ where q takes now any positive value [23] (see further Example 3, Section 3.1). The latter walks are not hyperuniform except if $q = d-1$ or $q = d/2-1$. In addition, the pdf of the endpoint distance was obtained for two–step walks with a beta step length distribution which depends now on two different scale parameters $q+s$ and q, where s is any positive integer [23].

1.4 General hyperuniform random walks

The focus of the present paper is on hyperuniform two-step random walks with a fixed total step length of $S = 1$ and step lengths of L and $1-L$. Its aim is then to derive all continuous distributions of L which yield two–step hyperuniform random walks, i.e. with distance distributions given by eq. (1). As justified in section 2, the emphasis will be placed on symmetric distributions of L on $[0,1]$ for which the two steps play equivalent roles. Our approach is thus the converse of those which derive a distribution of the endpoint distance from some step length distribution, for instance from a Dirichlet distribution. The two parameters of the study are those which allow defining the pdf $p_{d,k}(r)$ eq. (1): first the dimension $d \geq 2$ of the walk space and second the dimension $k > d$ of the hyperspace whose actual range of variation will be shown to depend on d. Hyperuniform random walks in $\mathbb{R}^d$ of special interest are those which yield uniform distributions of the endpoint on the unit ball B_1^d. The unique continuous and symmetrical distributions of L associated to uniform two-step random walks will be shown to have simple polynomial forms for any $d \geq 3$.

2 Why focus on symmetric step length distributions?

Let $S^L := \sum_{i=1}^{n} \boldsymbol{U}_i L_i$, where $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ are independent and identically distributed random unit vectors with uniform distribution on the sphere of $\mathbb{R}^d$, we call S^L the random walk associated to L. The sum which gives S^L is commutative. To phrase briefly the question which is considered in detail below, we notice that the $n!$ possible attributions of a set of lengths $(l_k, k = 1, \ldots, n)$ to the steps numbered $1, 2, .., n$ result in undistinguishable walks. Thus, each permutation should be given a probability of $1/n!$ so that all steps end up with length distributions independently of their arbitrary order. To obtain the latter distribution which is invariant under permutations, it suffices then to symmetrize the initial distribution of L if the latter is asymmetric. Profit will be taken from these symmetry considerations to restrict the study of step length distributions which yield hyperuniform random walks to those which are symmetric (Section 3).

2.1 Random walks of n steps

Let Σ be the set of permutations of $\{1,\ldots,n\}$. Given a random vector $\boldsymbol{L} \in \mathbb{R}^n$ and a permutation $\sigma \in \Sigma$, we will denote by L^σ the random vector $(L_{\sigma_1},\ldots,L_{\sigma_n})$. Furthermore, given a random permutation $\tilde{\sigma}$, that is a random vector taking values in Σ, we denote by $\tilde{L} := L^{\tilde{\sigma}}$ the vector obtained by this random permutation. We notice that $L^{\tilde{\sigma}} \,|\, (\tilde{\sigma} = \sigma) \overset{d}{=} L^\sigma$ (where $\overset{d}{=}$ means equality in distribution).

In particular, we will denote by L^* the random vector associated to $\tilde{\sigma}$ whose distribution on Σ is uniform and we will call L^* the *symmetrized* version of L.

Remark 1 It is easily seen that

1. L^* is random vector invariant under permutations distribution;
2. for all $\sigma \in \Sigma$, we have $(L^\sigma)^* \overset{d}{=} L^*$ and this implies that $\left(L^{\tilde{\sigma}}\right)^* \overset{d}{=} L^*$ for all random permutations $\tilde{\sigma}$.

If $\boldsymbol{L}$ is a continuous random vector, then the distribution of $\tilde{L}$ is obtained from

$$f_{\tilde{L}} = \sum_{\sigma \in \Sigma} f_{L^\sigma} \pi(\sigma)$$

where π is the distribution of $\tilde{\sigma}$ and $f_{\tilde{L}}$ and f_{L^σ} are, respectively the distributions of $\tilde{L}$ and L^σ.

Example 1 Let $L \sim D\left(q_1, q_2, q, \ldots, q\right)$, where $q_1 = q + p_1, q_2 = q + p_2, p_1, p_2 \in \mathbb{N}, p_1 + p_2 \geq 2$ and $q_i = q$, $i = 3, \ldots, n$. If $\tilde{\sigma}$ is a random permutation with distribution π, then

$$\begin{aligned}
f_{\tilde{L}}(\boldsymbol{\ell}) &= \frac{\Gamma(nq)\,(nq)_{p_1+p_2}}{\Gamma(q)^n\,(q)_{p_1}\,(q)_{p_2}} \sum_{\sigma\in\Sigma} \pi(\sigma) \prod_{h=1}^{n} \ell_h^{q_{\sigma(h)}-1} \\
&= \frac{\Gamma(nq)\,(nq)_{p_1+p_2}}{\Gamma(q)^n\,(q)_{p_1}\,(q)_{p_2}} \sum_{i\neq j} \sum_{\sigma\in\Sigma_{ij}} \pi(\sigma) \left(\prod_{h\notin\{i,j\}}^{n} \ell_h^{q-1} \right) \ell_i^{q+p_1} \ell_j^{q+p_2} \\
&= \frac{\Gamma(nq)}{\Gamma(q)^n} \prod_{h=1}^{n} \ell_h^{q-1} \frac{(nq)_{p_1+p_2}}{(q)_{p_1}\,(q)_{p_2}} \sum_{i\neq j} \pi_{ij} \ell_i^{p_1} \ell_j^{p_2}
\end{aligned}$$

where $\Sigma_{ij} := \{\sigma \in \Sigma : \sigma(i) = 1 \& \sigma(j) = 2\}$ is a partition of Σ and $\pi_{ij} = \pi(\Sigma_{ij})$.

In particular, if $\tilde{\sigma}$ is uniform, the density is

$$f_{\tilde{L}}(\boldsymbol{\ell}) = \frac{\Gamma(nq)(n-3)!\,(nq)_{p_1+p_2}}{n!\Gamma(q)^n\,(q)_{p_1}\,(q)_{p_2}} \prod_{h=1}^{n} \ell_h^{q-1} \sum_{i\neq j} \ell_i^{p_1} \ell_j^{p_2}$$

and this is a symmetric density, which is not a Dirichlet density, but is however a mixture of Dirichlet densities. The symmetric Dirichlet distribution is obtained only if there exists $(\pi_{ij})_{i\neq j}^{n}$ non–negative with the additional condition:

$$\frac{(nq)_{p_1+p_2}}{(q)_{p_1}\,(q)_{p_2}} \sum_{i\neq j} \pi_{ij} \ell_i^{p_1} \ell_j^{p_2} \equiv 1.$$

Dirichlet distributions are therefore not closed with respect to the symmetrization except in the particular case of Example 2 (next section).

It is easily seen that $S^L \stackrel{d}{=} S^{L^\sigma}$ for all $\sigma \in \Sigma$ and this implies that, if $\tilde{\sigma}$ is a Σ–valued random variable independent of S, then $S^L \stackrel{d}{=} S^{L^{\sigma}} \stackrel{d}{=} S^*$ where $S^* := S^{L^*}$ is the random walk whose step length distribution is symmetric. This means that the distribution of the endpoint of a random walk is invariant with respect to a random permutation of the components of the step–length random vector. Thus, the common endpoint distribution is that of the random walk whose step length distribution is symmetric.

These remarks allow us to partition the set of n–dimensional random vectors, denoted by $\mathscr{V}_n$, in classes of equivalent vectors with respect to the following definition

Definition 1 Given $X, Y \in \mathscr{V}_n$, X is in relation with Y, $X \sim Y$, if there exist two random permutations $\tilde{\sigma}_X$ and $\tilde{\sigma}_Y$ such that $X^{\tilde{\sigma}_X} \stackrel{d}{=} Y^{\tilde{\sigma}_Y}$.

By 2. of Remark 1, it follows that

Lemma 1 *Given $X, Y, Z \in \mathscr{V}_n$. If $X \sim Y$ and $Y \sim Z$, then $X^* = Z^*$ and $\sim$ is an equivalence relation on $\mathscr{V}_n$.*

The above have some important consequences:

1. The set of random vectors V_n can be partitioned into equivalence classes. The distributions of all random vectors which belong to the same class have associated symmetrized versions which are all equivalent. Any class contains one and only one permutation invariant distribution which is taken as the one representing the considered class.
2. If X and Y are in the same class, then $S^X \stackrel{d}{\sim} S^Y$

To find the distribution of the endpoint of any random walk of a given class, it suffices thus to the study the random walk whose step length distribution is the representative permutation invariant distribution of that class.

This fact can be stated as follows too: we define the random walk application at step n on $\mathscr{V}_n$ as

Definition 2 A random walk is an application $S : \mathscr{V}_n \longrightarrow \mathscr{V}_1$ such that $S(L) = S^L$.

From the point of view of distributions, it is equivalent to define S as above or to define it on the quotient space.

2.2 Two–step random walks

We consider now two–step random walks whose sum of step lengths is fixed and taken for convenience as equal to 1. The above definitions are thus simplified. Indeed, if we denote by L the length of the first step, then the distribution of the step length vector is symmetric if $L \overset{d}{=} 1 - L$. Furthermore, the space of permutation is given by $\Sigma = \{(1,2),(2,1)\}$ and the distribution of a random permutation $\tilde{\sigma}$ is completely identified by $p := \mathbb{P}\left(\tilde{\sigma} = (1,2)\right)$, then

$$f_{L^{\tilde{\sigma}}}(\ell) = pf_L(\ell) + (1-p)f_{1-L}(\ell) = pf_L(\ell) + (1-p)f_L(1-\ell)$$

and the symmetrized version has density: $f_{L^*}(\ell) = \left(f_L(\ell) + f_L\left(1-\ell\right)\right)/2$.

Example 2 The distribution of L is now taken to be an asymmetric Dirichlet distribution, that is $L \sim D\left(q+s, q\right), \, q, s > 0$ (actually a beta distribution), and we seek for the values of the parameters for which the associated symmetric distribution is again a Dirichlet distribution. As expected, the symmetrized distribution is the sole symmetric distribution which belongs to the family of mixtures considered above. It suffices therefore to look for the existence of a Dirichlet distribution only in the case where $p = 1/2$. Then it becomes

$$\begin{aligned} f^*(\ell) &= \frac{1}{2B\left(q+s,s\right)} \left[\ell^{q+s-1}(1-\ell)^{q-1} + \ell^{q-1}(1-\ell)^{q+s-1}\right] \\ &= \frac{1}{2B\left(q+s,s\right)} \, \ell^{q-1}(1-\ell)^{q-1}\left[\ell^s + (1-\ell)^s\right]. \end{aligned}$$

The distribution f^* is a Dirichlet distribution if and only if $\ell^s + (1-\ell)^s \equiv c$, where c is a constant, a condition which holds if and only if $s = 1$. When s differs from 1, the distribution f^* is a mixture of beta distributions.

The latter example, which is a particular case of Example 1, shows that a two–step random walk with an asymmetric Dirichlet distribution of step lengths, $D\left(q_1, q_2\right)$, is equivalent to a random walk with a symmetric step length Dirichlet distribution, $D\left(q, q\right)$, with $q = \min\left(q_1, q_2\right)$ if and only if $|q_1 - q_2| = 1$ [23].

3 Two-step random walks: a simple geometrical approach

Using a simple geometrical approach, we derive below the cumulative distribution function and the pdf of the endpoint distance of a constrained two–step random walk from its step length distribution. The case where the latter distribution is continuous with an associated density is more particularly considered (Corollary 1).

Let $\boldsymbol{S}_d(L)$ be the random vector

$$\boldsymbol{S}_d(L) = L\boldsymbol{U}_1 + (1-L)\boldsymbol{U}_2 \tag{3}$$

where $\boldsymbol{U}_1$, $\boldsymbol{U}_2$, are independent and identically distributed (i.i.d.) random unit vectors with uniform distribution on the sphere in $\mathbb{R}^d$ and L is a random variable (r.v.) whose support is included in $[0, 1]$.

We denote by $cap^d(\theta)$ the hyperspherical cap of geodesic radius θ on the boundary of B_1^d, and by $A_{cap}^d(\theta)$ its surface measure (see [25] for a simple expression of $A_{cap}^d(\theta)$).

In the next theorem we give a simple expression (4) to derive the distribution of $\boldsymbol{S}_d(L)$. The inner integral in (4), for which $L = l$, can be obtained for $n = 2$ from an integral given by Watson for a random walk of n a priori unequal stretches in a d-dimensional space ([37], top of page 421). It depends only on the measure of the intersection of a sphere with a ball in $\mathbb{R}^d$, that is the measure $A_{cap}^d(\theta)$. For the sake of clarity, we prove it here by a simple geometrical argument which relies on the isotropy of the considered random walk.

Theorem 1 *For $r \in [0,1]$, let $g_{d,L}(r)$ denote the probability that $\boldsymbol{S}_d(L)$ falls in the ball B_r^d, then*

$$g_{d,L}(r) = \mathbb{P}\left(\boldsymbol{S}_d(L) \in B_r^d\right) = C_d \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} \left[\int_0^{\theta(r,\ell)} \sin^{d-2}(\varphi)\, d\varphi\right] d\mathbb{P}_L(\ell) \tag{4}$$

where $C_d = \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)}$, $\mathbb{P}_L$ is the probability distribution of L and $\theta(r,\ell)$ satisfies

$$\frac{1+\cos\left(\theta(r,\ell)\right)}{2} = \frac{1-r^2}{4\ell(1-\ell)}. \tag{5}$$

Proof The probability that $\boldsymbol{S}_d(L)$ falls in B_r^d , is

$$\begin{aligned} g_{d,L}(r) = \mathbb{P}\left(\boldsymbol{S}_d(L) \in B_r^d\right) &= \mathbb{E}\left[\mathbb{P}\left(L\boldsymbol{U}_1 + (1-L)\boldsymbol{U}_2 \in B_r^d | L, \boldsymbol{U}_1\right)\right] \\ &= \mathbb{E}\left[\mathbb{P}\left(L\boldsymbol{e}_1 + (1-L)\,\boldsymbol{U}_2 \in B_r^d | L\right)\right] \end{aligned} \tag{6}$$

where $\boldsymbol{e}_1$ is the first vector of the standard base of $\mathbb{R}^d$ and the last equality follows from isotropy of the uniform distribution.

If the range of L is not in the interval $[(1-r))/2,(1+r)/2]$, then the conditional probability that the walk falls inside the ball of radius r, that is $\mathbb{P}\left(L\boldsymbol{e}_1+(1-L)\boldsymbol{U}_2\in B_r^d|L\right)$, is zero. Therefore

$$\begin{aligned}\mathbb{P}\left(L\boldsymbol{e}_1+\boldsymbol{U}_2(1-L)\in B_r^d|L\right)&=\mathbb{P}\left(\boldsymbol{U}_2\in cap\left(\theta\left(r,L\right)\right)|L\right)\mathbf{1}_{\left[\frac{1-r}{2},\frac{1+r}{2}\right]}(L)\\&=\frac{A_{cap}^d(\theta\left(r,L\right))}{A_{cap}^d\left(\pi\right)}\mathbf{1}_{\left[\frac{1-r}{2},\frac{1+r}{2}\right]}(L).\end{aligned}$$

Recalling that $A_{cap}^d(\theta)=\frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)}\int_0^\theta\sin^{d-2}\left(\varphi\right)d\varphi$ (see [25]), we get:

$$\mathbb{P}\left(L\boldsymbol{e}_1+\boldsymbol{U}_2(1-L)\in B_r^d|L\right)=\frac{\Gamma\left(d/2\right)}{\sqrt{\pi}\Gamma\left((d-1)/2\right)}\left(\int_0^{\theta(r,L)}\sin^{d-2}\left(\varphi\right)\,\mathrm{d}\varphi\right)\mathbf{1}_{\left[\frac{1-r}{2},\frac{1+r}{2}\right]}(L).$$

and finally, we obtain

$$g_{d,L}(r)=\frac{\Gamma\left(d/2\right)}{\sqrt{\pi}\Gamma\left((d-1)/2\right)}\mathbb{E}\left[\int_0^{\theta(r,L)}\sin^{d-2}\left(\varphi\right)\,\mathrm{d}\varphi\right]=C_d\int_{\frac{1-r}{2}}^{\frac{1+r}{2}}\left[\int_0^{\theta(r,\ell)}\sin^{d-2}(\varphi)\,\mathrm{d}\varphi\right]\mathrm{d}\mathbb{P}_L(\ell).$$

□

From now on, we will assume that the r.v. L has a continuous distribution in $[0,1]$ with density f_L. Applying Theorem 1 to this case, we obtain

Corollary 1 *If L has a density function f_L, for all $r\in[0,1]$, then*

$$g_{d,L}(r)=C_d\int_{\frac{1-r}{2}}^{\frac{1+r}{2}}f_L\left(\ell\right)\left[\int_0^{\theta(r,\ell)}\sin^{d-2}(\varphi)\,d\varphi\right]\,d\ell. \tag{7}$$

Furthermore, $\|\boldsymbol{S}_d(L)\|$ is a continuous random variable and it has the following density

$$g'_{d,L}(r)=2r\left(1-r^2\right)^{\frac{d-3}{2}}C_d\int_{\frac{1-r}{2}}^{\frac{1+r}{2}}\frac{f_L(\ell)}{\left(2\ell(1-\ell)\right)^{d-2}}\left[\left(4\ell(1-\ell)-(1-r^2)\right)\right]^{\frac{d-3}{2}}d\ell. \tag{8}$$

Proof Equation (7) is equation (4) assuming $\mathbb{P}_L$ to be continuous. The distribution function of $\|S_d(L)\|$ is simply $g_{d,L}$, which by hypothesis is a continuous function. Thus, the associated probability density is

$$\begin{aligned} g'_{d,L}(r) &= C_d \left\{ \frac{1}{2} f_L\left(\frac{1+r}{2}\right) \int_0^{\theta\left(r,\frac{1+r}{2}\right)} \sin^{d-2}(\varphi)\,\mathrm{d}\varphi + \right. \\ &\qquad + \frac{1}{2} f_L\left(\frac{1-r}{2}\right) \int_0^{\theta\left(r,\frac{1-r}{2}\right)} \sin^{d-2}(\varphi)\,\mathrm{d}\varphi \\ &\qquad \left. + \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} f_L(\ell)\,\frac{\partial\theta}{\partial r}(r,\ell) \sin^{d-2}\left(\theta(r,\ell)\right)\,\mathrm{d}\ell \right\} \\ &= C_d \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} f_L(\ell)\,\frac{\partial\theta}{\partial r}(r,\ell) \sin^{d-2}\left(\theta(r,\ell)\right)\,\mathrm{d}\ell \end{aligned} \tag{9}$$

where the last equality follows from (5) as $\theta\left(r,\frac{1\pm r}{2}\right)=0$.

Finally, from the derivative $\frac{\partial\theta}{\partial r}(r,\ell)$ and from $\sin(\theta) = \frac{\sqrt{1-r^2}\sqrt{4\ell(1-\ell)-(1-r^2)}}{2\ell(1-\ell)}$, we obtain (8). □

Example 3 Let $f_L(\ell) = \frac{1}{B(q,q)}\,[\ell(1-\ell)]^{q-1}$, that is $L \sim D(q,q)$. This is a symmetric continuous distribution and applying (8), we get

$$\begin{aligned} g'_{d,L}(r) &= \frac{r\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{2^{d-1}B(q,q)} \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} [\ell(1-\ell)]^{q-d+1}\left[\left(4\ell(1-\ell)-(1-r^2)\right)\right]^{\frac{d-3}{2}}\,\mathrm{d}\ell \\ &\overset{\ell=(1-y)/2}{=} \frac{r\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{2^{2q-d+2}B(q,q)} \int_{-r}^{r} \left(1-y^2\right)^{q-d+1}\left(r^2-y^2\right)^{\frac{d-3}{2}}\,\mathrm{d}y \\ &= \frac{r\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{2^{2q-d+1}B(q,q)} \int_0^{r} \left(1-y^2\right)^{q-d+1}\left(r^2-y^2\right)^{\frac{d-3}{2}}\,\mathrm{d}y \\ &\overset{z=y^2}{=} \frac{r\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{2^{2q-d+2}B(q,q)} \int_0^{r^2} z^{-1/2}\left(1-z\right)^{q-d+1}\left(r^2-z\right)^{\frac{d-3}{2}}\,\mathrm{d}z \\ &\overset{v=z/r^2}{=} \frac{r^{d-1}\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{2^{2q-d+2}B(q,q)} \int_0^{r^2} v^{-1/2}\left(1-r^2v\right)^{q-d+1}\left(1-v\right)^{\frac{d-3}{2}}\,\mathrm{d}v \\ &= \frac{r^{d-1}\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{2^{2(q+1)-d}} \frac{B\left(1/2,(d-1)/2\right)}{B(q,q)}\, {}_2F_1\left(d-q-1,1/2;d/2;r^2\right) \\ &= \frac{2^{d-1}r^{d-1}\left(1-r^2\right)^{\frac{d-3}{2}}}{B(q,1/2)}\, {}_2F_1\left(d-q-1,1/2;d/2;r^2\right) \end{aligned}$$

where the penultimate equality follows from 9.111 of [13]. The last expression was recently derived by a different method in [23].

Hereafter, we write:

$$h_1(r) := g'_{d,L}(r)/\left[r(1-r^2)^{\frac{d-3}{2}}\right] \text{ and } h_{k+1}(r) := \frac{h'_k(r)}{r},\ k \geq 1. \tag{10}$$

The following proposition emphasizes the relationship between the sequence h_m and the density f_L.

Proposition 1 *Given any f_L and $g'_{d,L}$ satisfying equation (8), if $d = 2m+3$, $m \geq 0$ and $r \in (0,1)$, we have*

$$h'_{m+1}(r) = C_{2m+3}m!2^{3m+1}\frac{f_L\left(\frac{1+r}{2}\right) + f_L\left(\frac{1-r}{2}\right)}{(1-r^2)^{2m+1}}. \tag{11}$$

In particular, in the symmetric case ($f_L = f_{1-L}$)

$$h'_{m+1}(r) = C_{2m+3}m!2^{3m+2}\frac{f_L\left(\frac{1+r}{2}\right)}{(1-r^2)^{2m+1}}. \tag{12}$$

Moreover, if $d = 2m+2$, $m \geq 1$, then we have:

$$h_{m+1}(r) = 2^{2m}C_{2m+2}(2m-1)!!\int_{-r}^{r}\frac{f_L\left(\frac{1+\ell}{2}\right)}{\sqrt{r^2-\ell^2}\,(1-\ell^2)^{2m}}\,d\ell. \tag{13}$$

Proof Applying (8) to the case $d = 2m+3$, we get

$$g'_{d,L}(r) = C_{2m+3}\int_{\frac{1-r}{2}}^{\frac{1+r}{2}}\frac{rf_L(\ell)}{2^{2m}[\ell(1-\ell)]^{2m+1}}\left[(1-r^2)\left(4\ell(1-\ell)-(1-r^2)\right)\right]^m \mathrm{d}\ell.$$

Dividing by $r(1-r^2)^m$, we obtain h_1 and taking the derivative with respect to r

$$h'_1(r) = \frac{mC_{2m+3}}{2^{2m-1}}\int_{\frac{1-r}{2}}^{\frac{1+r}{2}}\frac{rf_L(\ell)}{[\ell(1-\ell)]^{2m+1}}\left(4\ell(1-\ell)-(1-r^2)\right)^{m-1}\mathrm{d}\ell. \tag{14}$$

Then, obviously,

$$h_2(r) := h_1'(r)/r = \frac{mC_{2m+3}}{2^{2m-1}} \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} \frac{f_L(\ell)}{[\ell(1-\ell)]^{2m+1}} \left(4\ell(1-\ell) - (1-r^2)\right)^{m-1} \mathrm{d}\ell$$

and repeating the same procedure $m-1$ times, we get

$$h_{m+1}(r) = \frac{C_{2m+3}m!}{2^{m+1}} \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} \frac{f_L(\ell)}{[\ell(1-\ell)]^{2m+1}} \,\mathrm{d}\ell.$$

Finally, taking the derivative another time with respect to r, we get (11) while (12) is easily obtained from the symmetry condition.

If the dimension is even, $d = 2m+2$, we can proceed analogously. □

4 Two–step hyperuniform random walks

The previous theorems are first applied to hyperuniform random walks of two steps $HU_d(k)$ to determine the allowed range of the hyperspace dimension k given the walk space dimension d. Then the step length pdf's are explicitly derived as a function of k and d.

4.1 The allowed range of k for $HU_d(k)$ walks

In the following proposition, the maximum dimension k of the $HU_d(k)$ random walks of two steps is characterized.

Proposition 2 *If a two–step random walk is $HU_d(k)$, then $d+1 \le k \le 2d-1$.*

Proof Obviously, $k \ge d+1$, so we prove only that $k \le 2d-1$. Suppose, contrary to our claim, that there exists $k \ge 2d$, such that the random walk is $HU_d(k)$.

Then, by (1),

$$g'_{d,L}(r) = \frac{2\Gamma(k/2)}{\Gamma(d/2)\Gamma((k-d)/2)} r^{d-1}(1-r^2)^{\frac{k-d-2}{2}}$$

and

$$h_1(r) = \frac{2\Gamma(k/2)}{\Gamma(d/2)\Gamma((k-d)/2)} r^{d-2}(1-r^2)^{\frac{k-2d+1}{2}}$$

its derivative is

$$h'_1(r) = \frac{2\Gamma(k/2)}{\Gamma(d/2)\Gamma((k-d)/2)} r^{d-3}(1-r^2)^{\frac{k-2d-1}{2}} \left[d-2-(k-(d+1))r^2\right]$$

which is negative in the interval $(\sqrt{(d-2)/(k-(d+1))}, 1]$ if $k \geq 2d$. Furthermore, from (14), noticing that it holds in the even and odd cases, it follows that

$$h'_1(r) = \frac{2C_d}{2^{2d-3}} \int_{\frac{1-r}{2}}^{\frac{1+r}{2}} \frac{r f_L(\ell)}{[\ell(1-\ell)]^{d-2}} \left(4\ell(1-\ell)-(1-r^2)\right)^{m-1} \,\mathrm{d}\ell$$

then for all r in the above interval, there exists a set on which the integrand must be negative. This implies that f_L is negative on some set with non–null measure and this is impossible, because f_L is a density. □

4.2 The continuous and symmetric step length densities for $HU_d(k)$ walks

Theorem 2 *Given $d+1 \leq k \leq 2d-1$, the 2-step random walk $\boldsymbol{S}_d(L) = L\boldsymbol{U}_1 + (1-L)\boldsymbol{U}_2$ is a hyperuniform walk $HU_d(k)$ if and only if the symmetric density function $f_L(l)$ of L satisfies:*

$$f_L(l) = \frac{2^{d-1}}{B(d/2,(k-d)/2)} \left[l(1-l)\right]^{d-2} \times {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; 1-4l(1-l)\right). \tag{15}$$

Proof Let $\boldsymbol{S}_d(L) = L\boldsymbol{U}_1 + (1-L)\boldsymbol{U}_2$ and we set

$$R^2 := \boldsymbol{S}_d(L)\cdot\boldsymbol{S}_d(L) = 1 - 2L\,(1-L)\,(1-\boldsymbol{U}_1\cdot\boldsymbol{U}_2)\,.$$

Also let Θ be such that $\boldsymbol{U}_1\cdot\boldsymbol{U}_2 = \cos\Theta$ and

$$\begin{cases} V = 1 - R^2 = 2L\,(1-L)\,(1-\cos\Theta) \\ Y = 4L(1-L) \\ Z = \dfrac{1-\cos\Theta}{2} = \sin^2\left(\dfrac{\Theta}{2}\right) \end{cases}.$$

With this notation

$$V = YZ \tag{16}$$

where Y and Z are independent random variables, being functions of independent random variables. The problem of finding the distribution of Y given the distributions of V and Z, where V, Y, Z are univariate continuous positive random variables, belongs to the class of inverse problems of random scaling ([3] and references therein). General considerations on the case where Z has a beta distribution with positive parameters are given in [17] and in the references therein.

Assuming that $\boldsymbol{S}_d(L)$ is $HU_d(k)$, then

$$V \sim Beta\left(\frac{k-d}{2}, \frac{d}{2}\right) \text{ and } Z \sim Beta\left(\frac{d-1}{2}, \frac{d-1}{2}\right).$$

where the first equality in distribution follows from (1) (see too [24]) and the second one from the distribution of the polar angle Θ, which is, for $\theta \in [0,\pi]$:

$$p_\Theta(\theta) = \frac{\Gamma(d/2)}{\sqrt{\pi}\Gamma((d-1)/2)}\sin^{d-2}(\theta) = \frac{2^{d-2}\Gamma(d/2)}{\sqrt{\pi}\Gamma((d-1)/2)}\sin^{d-2}(\theta/2)\cos^{d-2}(\theta/2)$$

which yields immediately the distribution of Z.

Now, we use the Mellin transform to find the distribution of Y. The Mellin transform of $Beta\,(\alpha,\beta)$ with density f is

$$f^*(s) := \int_0^1 z^{s-1} f(z)\,\mathrm{d}z = \frac{\Gamma\,(\alpha+s-1)\,\Gamma\,(\alpha+\beta)}{\Gamma\,(\alpha)\,\Gamma\,(\alpha+\beta+s-1)} \tag{17}$$

Then, we have

$$f_Z^*(s) = \frac{\Gamma(\frac{d-3}{2}+s)\Gamma(d-1)}{\Gamma(\frac{d-1}{2})\Gamma(d-2+s)} \text{ and } f_V^*(s) = \frac{\Gamma(\frac{k-d-2}{2}+s)\Gamma(k/2)}{\Gamma(\frac{k-d}{2})\Gamma(\frac{k-2}{2}+s)}.$$

By the independence of Y and Z, the Mellin transform of Y is

$$f_Y^*(s) = \frac{f_V^*(s)}{f_Z^*(s)} = \frac{\Gamma((d-1)/2)\Gamma(k/2)}{\Gamma(d-1)\Gamma((k-d)/2)} \times \frac{\Gamma(\frac{k-d-2}{2}+s)\Gamma(d-2+s)}{\Gamma(\frac{k-2}{2}+s)\Gamma(\frac{d-3}{2}+s)}. \tag{18}$$

By the Mellin inversion theorem, the density of Y is given by

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} f^*(s)x^{-s}\,\mathrm{d}s \\
&= \frac{1}{2\pi i}\,\frac{\Gamma((d-1)/2)\Gamma(k/2)}{\Gamma(d-1)\Gamma((k-d)/2)}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{\Gamma(\frac{k-d-2}{2}+s)\Gamma(d-2+s)}{\Gamma(\frac{k-2}{2}+s)\Gamma(\frac{d-3}{2}+s)}x^{-s}\,\mathrm{d}s \\
&= \frac{\Gamma((d-1)/2)\Gamma(k/2)}{\Gamma(d-1)\Gamma((k-d)/2)}\,G_{2,2}^{2,0}\left(y\,\middle|\,\begin{matrix}\frac{k-2}{2}\ ,\ \frac{d-3}{2}\\[2ex] \frac{k-d-2}{2}\ ,\ d-2\end{matrix}\right)
\end{aligned}$$

where the last equality follows from the definition of Meijer's G function (see 9.3.1 of [13]). Furthermore, according to the following identity, for $|z|<1$ (see [27])

$$\begin{aligned}
&G_{2,2}^{2,0}\left(z\,\middle|\,\begin{matrix}\alpha_1+\beta_1-1\ ,\ \alpha_2+\beta_2-1\\[2ex] \alpha_1-1\quad ,\quad \alpha_2-1\end{matrix}\right) = \\
&\qquad = \frac{z^{\alpha_2-1}(1-z)^{\beta_1+\beta_2-1}}{\Gamma(\beta_1+\beta_2)}\ {}_2F_1\left(\alpha_2+\beta_2-\alpha_1,\beta_1;\beta_1+\beta_2;1-z\right)
\end{aligned}$$

we obtain for $y \in [0,1]$

$$f_Y(y) = \frac{\Gamma((d-1)/2)\Gamma(k/2)}{\Gamma(d-1)\Gamma((k-d)/2)} \frac{y^{d-2}(1-y)^{-1/2}}{\sqrt{\pi}} \, {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; 1-y\right).$$

Applying the Legendre duplication formula 8.335.1 of [13] to $\Gamma(d-1)$, we get

$$f_Y(y) = \frac{2^{2-d}}{B(d/2,(k-d)/2)} \, y^{d-2}(1-y)^{-1/2} \, {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; 1-y\right). \tag{19}$$

Recalling that

$$\Pr(Y \le y) = \Pr\left(L \le \frac{1-\sqrt{1-y}}{2}\right) + \Pr\left(L > \frac{1+\sqrt{1-y}}{2}\right)$$

then taking the derivative, we have that f_L must satisfies the following condition

$$f_Y(y) = \frac{4}{\sqrt{1-y}} \left\{ f_L\left(\frac{1-\sqrt{1-y}}{2})\right) + f_L\left(\frac{1+\sqrt{1-y}}{2}\right)\right\}.$$

Finally, from (19), we obtain for $l \in [0,1]$

$$f_L(l) = \frac{2^{d-1}}{B(d/2,(k-d)/2)} \, [l(1-l)]^{d-2} \, {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; 1-4l(1-l)\right).$$

To prove the sufficient condition in the relevant case, i.e. in the symmetric case (section 2), we assume that the step length distribution is given by eq.

(15). Then applying eq. (8) we get:

$$
\begin{aligned}
g'_{d,L}(r) &= \frac{4r\left(1-r^2\right)^{\frac{d-3}{2}} C_d}{B\left(\frac{d}{2}, \frac{k-d}{2}\right)} \times \\
&\times \int_{\frac{1}{2}}^{\frac{1+r}{2}} {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; 1-4l(1-l)\right)\left[\left(4\ell(1-\ell)-(1-r^2)\right)\right]^{\frac{d-3}{2}} \mathrm{d}\ell \\
&\overset{\ell,=(1+y)/2}{=} \frac{2r\left(1-r^2\right)^{\frac{d-3}{2}} \Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(\frac{k-d}{2}\right)} \int_0^r \left(r^2-y^2\right)^{\frac{d-3}{2}} {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; y^2\right) \mathrm{d}y \\
&\overset{y=r\sqrt{w}}{=} \frac{2r^{d-1}\left(1-r^2\right)^{\frac{d-3}{2}} \Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(\frac{k-d}{2}\right)} \int_0^1 w^{-1/2}\left(1-w\right)^{\frac{d-3}{2}} {}_2F_1\left(\frac{2d-k-1}{2}, \frac{d}{2}; \frac{1}{2}; r^2 w\right) \mathrm{d}w \\
&= \frac{2r^{d-1}\left(1-r^2\right)^{\frac{d-3}{2}} \Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(\frac{k-d}{2}\right)} B\left(\frac{1}{2}, \frac{d-1}{2}\right)\left(1-r^2\right)^{-\frac{2d-k-1}{2}} \\
&= \frac{2\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{k-d}{2}\right)} r^{d-1}(1-r^2)^{\frac{k-d-2}{2}}
\end{aligned}
$$

where we used formula 7.512.6 in [13]. The latter density is indeed seen to be the density of the endpoint distance of the hyperuniform random walk $HU_d(k)$ (eq. 1). □

From now on, we denote by $f_{d,k}(l)$ the symmetric density of the step length obtained in (15). When the first argument of the hypergeometric function of eq. (15), ${}_2F_1\left((2d-k-1)/2, d/2; 1/2; 1-4l(1-l)\right)$, is made equal to zero, that is $k = 2d-1$, then the step length density reduces to a beta distribution, namely:

$$
f_{d,2d-1}(l) = \frac{2^{d-1}\left[l(1-l)\right]^{d-2}}{B(\frac{d}{2}, \frac{d-1}{2})} = \frac{\left[l(1-l)\right]^{d-2}}{B(d-1,d-1)}. \tag{20}
$$

The classical transformation (eq. 9.131.1 of [13]) applied to ${}_2F_1((2d-k-1)/2, d/2; 1/2; 1-4l(1-l))$ gives $(4l(1-l))^{(k-3d+2)/2}\ {}_2F_1((k-2d+2)/2, (1-d)/2; 1/2; 1-4l(1-l))$. The first argument of the transformed hypergeometric function is zero for $k = 2d - 2$ and the step length becomes again a beta random variable:

$$f_{d,2d-2}(l) = \frac{[l(1-l)]^{\frac{d-4}{2}}}{2B(\frac{d}{2}, \frac{d-2}{2})} = \frac{[l(1-l)]^{\frac{d-4}{2}}}{B(\frac{d}{2}-1, \frac{d}{2}-1)}. \tag{21}$$

Distribution (21) is a $Beta\,(d/2-1, d/2-1)$. Both results agree with those on hyperuniform random walks given in [21].

4.3 The random walk $HU_d(d+2)$ with a uniform distribution on B_1^d

A transformation is applied now to the hypergeometric function of (15) to get a simpler explicit expression of the step length density for random walks $HU_d(d+2)$ whose endpoints are uniformly distributed on the ball B_1^d. We notice that it is necessary to assume $d \geq 3$ because, by Proposition 2, k can take a unique value, $k = 3$, for $d = 2$. As a uniform random walk requires $k = d+2 = 4$ for $d = 2$, the latter cannot be uniform. In sum, there exist no uniform two–step random walk with a continuous step length distribution in $\mathbb{R}^2$.

Inserting the following relation (eq. 15.8.27 of [28]):

$$\begin{aligned}{}_2F_1\left(\alpha, \beta; \frac{1}{2}; z\right) &= \frac{\Gamma(\alpha+1/2)\Gamma(\beta+1/2)}{2\sqrt{\pi}\Gamma(\alpha+\beta+1/2)} \times \\ &\times \left({}_2F_1(2\alpha, 2\beta; \alpha+\beta+1/2; \frac{1-\sqrt{z}}{2}) + {}_2F_1(2\alpha, 2\beta; \alpha+\beta+1/2; \frac{1+\sqrt{z}}{2})\right)\end{aligned}$$

into (15) yields another representation of the density:

$$\begin{aligned}f_{d,k}(l) = C(d,k)\,[l(1-l)]^{d-2} \Bigg\{ &{}_2F_1\left(2d-1-k, d; \frac{3d-k}{2}; l\right) + \\ &+ {}_2F_1\left(2d-1-k, d; \frac{3d-k}{2}; 1-l\right)\Bigg\}\end{aligned} \tag{22}$$

where

$$C(d,k)=\frac{2^{d-2}\Gamma\left(\frac{2d-k}{2}\right)\Gamma\left(\frac{d+1}{2}\right)}{\sqrt{\pi}B\left(\frac{d}{2},\frac{k-d}{2}\right)\Gamma\left(\frac{3d-k}{2}\right)}$$

In particular, for $k=d+2$, applying the duplication formula to the latter constant, we obtain

$$f_{d,d+2}(l)=\frac{d(d-1)}{2(d-2)}\left[l(1-l)\right]^{d-2}\left\{\ {}_2F_1\left(d-3,d;d-1;l\right)+\ {}_2F_1\left(d-3,d;d-1;1-l\right)\right\}$$

moreover, applying 9.131.1 of [13], i.e.

$${}_2F_1(d-3,d;d-1;x)=(1-x)^{2-d}\,{}_2F_1(2,-1;d-1;x)=2(1-x)^{2-d}((d-1)/2-x)/(d-1),$$

we get a simple polynomial expression for the density:

$$f_{d,d+2}(l)=\frac{d}{d-2}\times\left\{(1-l)^{d-2}\left(l+\frac{d-3}{2}\right)+l^{d-2}\left(1-l+\frac{d-3}{2}\right)\right\}. \quad (23)$$

where the symmetry between l and $1-l$ has been purposely emphasized in the writing of (23). After expanding the products in (23), we get

$$f_{d,d+2}(l)=\begin{cases}\frac{d}{2(d-2)}\left(\sum_{n=0}^{d-4}(-1)^n(d-3-n)\binom{d-1}{n}l^n\right) & \text{if } d \text{ is even}\\ \frac{d}{d-2}\sum_{n=0}^{d-1}b_n^{(d)}l^n & \text{if } d \text{ is odd}\end{cases}$$

where

$$b_n^{(d)}=\begin{cases}(-1)^n\dfrac{d-3-n}{2}\dbinom{d-1}{n} & (0\le n\le d-4)\quad(d\ge 5)\\ (-1)^n(d-3-n)\dbinom{d-1}{n} & (d-3\le n\le d-1)\quad(d\ge 3)\end{cases}.$$

The moments obtained from (23) are linear combinations of beta functions. Indeed:

$$E(L^n)=\frac{d}{d-2}\times\int_0^1 l^n\left\{(1-l)^{d-2}\left(l+\frac{d-3}{2}\right)+l^{d-2}\left(1-l+\frac{d-3}{2}\right)\right\}dl.$$

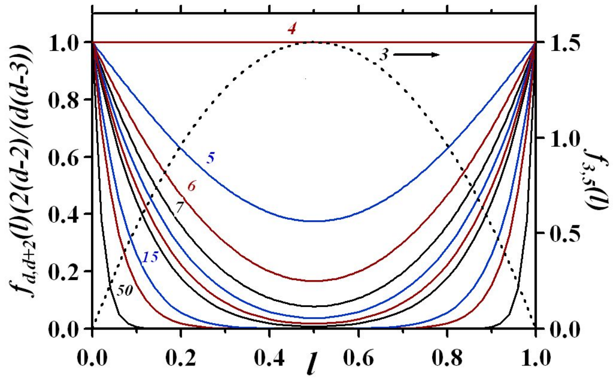

Fig. 1 Step-length pdf's $f_{d,d+2}(l)$ [see eq. (23)] which yield two–step random walks whose endpoints are uniformly distributed on the unit ball B_1^d: 1) left axis, solid lines: rescaled pdf's , $f_{d,d+2}(l)/f_{d,d+2}(0)$, with $d = 4(f_{4,6}(l) = 1), d = 5, 6, 7, 8, 9, 10, 15, 20, 50$; 2) right axis, dotted line, $d = 3$, $f_{3,5}(l) = 6l(1 - l)$

Therefore:

$$\begin{aligned}E(L^n) &= \frac{d}{d-2} \times \{B(n+2, d-1) + B(n+d-1, 2)\} + \\ &\qquad + \frac{d(d-3)}{2(d-2)} \times \left\{ B(n+1, d-1) + \frac{1}{n+d-1} \right\} \\ &= \frac{d\left(d^2 + (n-3)d - 3n + 2\right)}{2(d-2)(n+d-1)(n+d)} + \frac{(n+d-2)\Gamma(d+1)\Gamma(n+1)}{2(d-2)\Gamma(n+d+1)}\end{aligned}$$

Furthermore, using the asymptotic relation for $d \to \infty$:

$$\frac{\Gamma(d+1)}{\Gamma(d+n+1)} \sim d^{-n} \left(1 - \frac{n(n+1)}{2d} + \cdots\right)$$

we have,

$$\lim_{d\to\infty} E(L^n) = \frac{1}{2}$$

The trend of L to concentrate around 0 and 1 when d becomes larger and larger is clearly shown by the densities $f_{d,d+2}(l)$ in figure 1. The endpoints of such uniform two–step random walks tend to concentrate near the surface of the unit ball as expected in high dimension.

4.4 Construction of arbitrary step length pdf's for hyperuniform two–step random walks

An infinity of asymmetric step length pdf's $f(l)$ can be constructed from a given symmetric step length pdf $f_s(l)$ ($f_s(l) = f_s(1-l)$) by adding to it an antisymmetric function $f_a(l)$ ($f_a(l) = -f_a(1-l)$). The antisymmetric function is arbitrary but chosen such that the resulting density $f(l)$ is non-negative for $l \in [0,1]$. The latter can be written as:

$$f(l) = f_s(l) + f_a(l) = \frac{f(l) + f(1-l)}{2} + \frac{f(l) - f(1-l)}{2} \quad (l \in [0,1])$$

This is illustrated with the following example for a uniform walk on B_1^3 with $d = 3$ and $k = 5$. The symmetric step length distribution is (23), $f_s(l) = f_{3,5}(l) = 6l(1-l)$ and the antisymmetric function is chosen to be:

$$f_a(l) = \begin{cases} 6l^2 - 3l & l \leq 1/2 \\ 3(1-l)(2l-1) & l \geq 1/2 \end{cases}$$

Then

$$f(l) = \begin{cases} 3l & l \leq 1/2 \\ 3(1-l)(4l-1) & l \geq 1/2 \end{cases}$$

An asymmetric pdf of step length constructed by the methods described above gives rise to a random walk which is equivalent to the walk obtained from the reference symmetric density. The degree of freedom in the construction of such asymmetric distributions may be used to tailor a convenient density for Monte–Carlo simulations.

5 Conclusions

Two–step random walks in Euclidean space $\mathbb{R}^d$, with fixed sums of step lengths, taken as equal to 1, have been investigated with the aim of constructing in a simple way continuous step length distributions which yield hyperuniform random walks. For the latter walks, the endpoint positions are spread out as are the projections in the walk space of points uniformly distributed on the surface of the unit hypersphere in $\mathbb{R}^k$ with $k > d$. When $k = d + 2$, a hyperuniform random walk becomes uniform on the unit ball of $\mathbb{R}^d$. Two random walks, the first with an asymmetric distribution of step length and the second with a permutation invariant step length distribution associated with the latter, have identical endpoint distributions. Symmetric distributions are thus the true reference distributions of step length as the equivalence between the two steps cancels an eventual initial asymmetry. As shown in section 4.4, an infinity of continuous asymmetric step length pdfs can be constructed from a given symmetric pdf. We have derived the unique symmetric continuous distributions of step length on $[0, 1]$ which yield hyperuniform two-step random walks. We have proven that the latter walks exist only for $d + 1 \le k \le 2d - 1$. Interestingly, the derived step length distributions for the two largest possible values of $k, 2d - 2, 2d - 1$ are Dirichlet distributions which reduce to beta distributions in the case of two steps. The question naturally arises as to whether it is possible to construct permutation invariant continuous distributions of step lengths which yield hyperuniform walks for any space dimension and any number of steps. In any case, the step length distributions of uniform random walks cannot be of the Dirichlet type as soon as $d \geq 4$ for a number of steps at least equal to three [10,19,21,29]. Indeed, uniform three–step walks are obtained solely for the two following symmetric Dirichlet densities (table 2 of [21]) ($l_3 = 1 - l_1 - l_2$): $D(1,1,1)$ with $p\,(l_1, l_2) = 2$ for $d = 2$ and $D(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ with $p\,(l_1, l_2) = \frac{1}{2\pi\sqrt{l_1 l_2 l_3}}$ for $d = 3$.

Appendix

A. Multivariate Dirichlet distribution from gamma distributions

The probability density function $p_X(x)$ of a gamma random variable X , is [9]

$$p_X(x) = \frac{x^{\alpha-1}\exp\left(-x/\theta\right)}{\theta^{\alpha}\Gamma(\alpha)} \;\; (x > 0) \tag{A.1}$$

We denote it here as $X \sim \gamma(\alpha, \theta)$ where $\alpha > 0$ is the shape parameter and $\theta > 0$ the scale parameter while $\Gamma(\alpha)$ is the Euler gamma function. The characteristic function of X is $\phi_X(t) = E(e^{itX}) = \frac{1}{(1-i\theta t)^{\alpha}}$ [9].

A sum $G = \sum_{i=1}^{n} G_i$ of n independent gamma random variables, $G_i \sim \gamma(q_i, \theta)$ $(i = 1, \cdots, n)$, with identical scale parameter θ and a priori different shape parameters q_i $(i = 1, \cdots, n)$, is a gamma random variable $G \sim \gamma(n\bar{q}, \theta)$, where $n\bar{q}$ is the sum $\sum_{i=1}^{n} q_i$. This is for instance deduced from the characteristic function of G,

$$\phi_G(t) = E(e^{itG}) = \frac{1}{\prod_{k=1}^{n}(1 - it\theta)^{q_k}} = \frac{1}{(1 - it\theta)^{n\bar{q}}}$$

From the previous set of $n = m + 1$ independent gamma random variables, we define a random vector $\boldsymbol{L}_{(n)} = (L_1, L_2, \cdots, L_n)$ whose components are $L_i = G_i/G$. The distribution of $\boldsymbol{L}_{(n)}$ is then called a Dirichlet distribution with parameters $\boldsymbol{q_{(n)}} = (q_1, \cdots, q_n)$, $\boldsymbol{L_{(n)}} \sim D(\boldsymbol{q_{(n)}})$. Its pdf is shown to be ([9] p. 17)

$$\begin{cases} f(l_1, \cdots, l_m) = \left(\frac{\Gamma(n\bar{q})}{\prod_{i=1}^n \Gamma(q_i)}\right) \prod_{i=1}^n l_i^{q_i - 1} \\ \\ l_n = 1 - \sum_{i=1}^m l_i, l_i > 0, i = 1, \cdots, n \end{cases} \tag{A.2}$$

When the shape parameters are all equal to q, i. e. when $G_i \sim \gamma(q, \theta)$ $(i = 1, \cdots, n)$, the pdf (A.2) becomes invariant under permutation, $f(l_1, \cdots, l_m) = \left(\frac{\Gamma(nq)}{\Gamma(q)^n}\right) \left\{\prod_{i=1}^n l_i\right\}^{q-1}$. Finally, when the G_i's are exponentially distributed, $q = 1$, the random vector $\boldsymbol{L}_{(n)}$ is uniformly distributed over the unit $(n-1)$ simplex as $f(l_1, \cdots, l_m) = m!$.

B. Hypergeometrical differential equation characterization

We recall that $F(l) := {}_2F_1(a, b; c; l)$ is the solution of the following differential equation:

$$l(1-l)F''(l) + [c - (a+b+1)l]F'(l) - abF(l) = 0$$

so, letting ${}_2F_1\left(2d-1-k, d; \frac{3d-k}{2}; l\right) = F(l)$, we have

$$l(1-l)F''(l) + \left[\frac{3d-k}{2} - (3d-k)l\right] F'(l) - d(2d-1-k)F(l) = 0$$

and similarly

$$l(1-l)F''(1-l) + \left[\frac{3d-k}{2} - (3d-k)(1-l)\right] F'(1-l) - d(2d-1-k)F(1-l) = 0$$

that is

$$l(1-l)F''(1-l) - \left[\frac{3d-k}{2} - (3d-k)l\right] F'(1-l) - d(2d-1-k)F(1-l) = 0$$

Let $G(l) = \frac{f_{d,k}(l)}{[l(1-l)]^{d-2}}$, by (22), we get

$$G(l) = C\left\{F(l) + F(1-l)\right\}.$$

Taking the derivative two times, we obtain

$$G'(l) = C\left\{F'(l) - F'(1-l)\right\} \quad \text{and} \quad G''(l) = C\left\{F''(l) + F''(1-l)\right\}$$

and substituting

$$l(1-l)G''(l)+\left[\frac{3d-k}{2}-(3d-k)l\right]G'(l)-d(2d-1-k)G(l)=0 \tag{B.1}$$

It follows that G is also a solution of a hypergeometrical differential equation. For basic results on these differential equations see [1], Table 15.5.

For $-1<r<1$ the change of variable $l=\frac{1+r}{2}$ in (B.1) gives

$$(1-r^2)F''(r)-(3d-k)rF'(r)-d(2d-1-k)F(r)=0$$

where $F(r)=G(\frac{1+r}{2})$. This is the the Gegenbauer equation which has a unique symmetric solution such that: $F(0)=C$ and $F'(0)=0$. It is $F(r)=C\times\ {}_2F_1(\frac{2d-k-1}{2},\frac{d}{2},\frac{1}{2},r^2)$, see [6], pag. 65, the density (15) is easily retrieved.

In the following proposition we prove that the only hyperuniform random walks $HU_d(k)$ which are at the same time symmetric Dirichlet random walks are obtained for $k=2d-1$ and $k=2d-2$ as shown in eqs (20) and (21).

Proposition 3 *Suppose that $f_L(l)=C\,[l(1-l)]^{\alpha-1}$, then $\boldsymbol{S}_d(L)$ is $HU_d(k)$ if and only if $k=2d-1$ and $\alpha=d-1$ or $k=2d-2$ and $\alpha=\frac{d}{2}-1$.*

Proof Suppose that $f_{d,k}(l)=C\,[l(1-l)]^{\alpha-1}$. Then $G(l)=C\,[l(1-l)]^{\alpha-d+1}$ and we have

$$G'(l)=(\alpha-d+1)\,[l(1-l)]^{\alpha-d}\,(1-2l)$$

$$G''(l)=(\alpha-d+1)\,[l(1-l)]^{\alpha-d-1}\,\{(\alpha-d)(1-2l)^2-2l+2l^2\}$$

So from (B.1) we obtain:

$$(\alpha-d+1)\{(\alpha-d)(1-2l)^2-2l+2l^2\}+\frac{3d-k}{2}(\alpha-d+1)(1-2l)^2-d(2d-1-k)(l-l^2)=0$$

which is of the form $Bl^2-Bl+A=0$ with

$$A=(\alpha-d+1)\{(\alpha-d)+\frac{3d-k}{2}\}$$

and

$$B=4(\alpha-d+1)(\alpha-d)+2(\alpha-d+1)+2(3d-k)(\alpha-d+1)+d(2d-1-k)$$

From the condition $A = B = 0$, we get the only solutions

$$k = 2d - 1,\ \alpha = d - 1 \text{ and } k = 2d - 2,\ \ \alpha = \frac{d}{2} - 1.$$

□

These solutions were obtained by a different method in [21].

References

1. Abramowitz, M., Stegun, I.A.: Handbook of mathematical functions with formulas, graphs, and mathematical tables. National Bureau of Standards Applied Mathematics Series, 55, (1964)
2. Aitchison, J.: The Statistical Analysis of Compositional Data. Chapman and Hall, London, (1986)
3. Balakrishnan, N., Stepanov, A.: On the use of bivariate Mellin transform in bivariate random scaling and some applications. Methodol. Comput. Appl. Probab. 16(1), 235–244, (2014)
4. Chandrasekhar, S.: Stochastic problems in physics and astronomy. Rev. Mod. Phys. 15, 1–89, (1943)
5. Codling, E. A., Plank, M. J., Benhamou, S.: Random walk models in biology. J. R. Soc. Interface 5, 813–834, (2008)
6. Dereziński, J.: Hypergeometric type functions and their symmetries. Ann. Henri Poincaré 15, 1569–1653, (2014)
7. Driver, K., Möller, M.: Quadratic and cubic transformations and zeros of hypergeometric polynomials. J. Comput. Appl. Math. 142, 411–417, (2002)
8. Dutka, J.: On the problem of random flights. Arch. Hist. Exact Sci. 32(3-4), 351–375, (1985)
9. Fang, K.T., Kotz, S., Ng, K.W.: Symmetric multivariate and related distributions. Monographs on Statistics and Applied Probability. Vol. 36, Chapman and Hall Ltd., London, (1990)

10. Franceschetti, M.: When a random walk of fixed length can lead uniformly anywhere inside a hypersphere. J. Stat. Phys. 127(4), 813–823, (2007)

11. García-Pelayo, R.: The random flight and the persistent random walk. In A. Skogseid and V. Fasano, editors, Statistical mechanics and random walks: principles, processes and applications, chapter The Random Flight and the Persistent Random Walk. Nova Science Publishers, (2011)

12. García-Pelayo, R.: Exact solutions for isotropic random flights in odd dimensions. J. Math. Phys. 53(10), 103504, 15pp, (2012)

13. Gradshteyn, I.S., Ryzhik, I.M.: Table of integrals, series, and products. Elsevier/Academic Press, Amsterdam, seventh edition, (2007)

14. De Gregorio, A.: A family of random walks with generalized Dirichlet steps. J. Math. Phys. 55(2), 02330, 17pp, (2014)

15. De Gregorio, A., Orsingher, E.: Flying randomly in $\mathbb{R}^d$ with Dirichlet displacements. Stochastic Process. Appl. 122(2), 676–713, (2011)

16. Hansen, M., Frejsel, A. M., Kim, J., Naselsky, P., Nesti, F.: Pearson's random walk in the space of the CMB phases: Evidence for parity asymmetry. Phys. Rev. D 83, 103509, (2011) (9 pages)

17. Hashorva, E., Pakes, A.G.: Tail asymptotics under beta random scaling. J. Math. Anal. Appl. 372, 496–514, (2010)

18. Kolesnik, A.D.: Probability Law For the Euclidean Distance Between Two Planar Random Flights. J. Stat. Phys. 154(4), 1124–1152, (2014)

19. Kolesnik, A.D.: Random motions at finite speed in higher dimensions. J. Stat. Phys. 131(6), 1039–1065, (2008)

20. Kolesnik, A.D.: The explicit probability distribution of a six-dimensional random flight. Theory Stoch. Process. 15(1), 33–39, (2009)

21. Le Caër, G.: A Pearson random walk with steps of uniform orientation and Dirichlet distributed lengths. J. Stat. Phys. 140(4), 728–751, (2010)

22. Le Caër, G.: A New Family of Solvable Pearson-Dirichlet Random Walks. J. Stat. Phys. 144(1), 23–45, (2011)

23. Le Caër, G.: Two–step Dirichlet random walks. Physica A (on line).

24. Letac, G., Piccioni, M.: Dirichlet random walks. e-print arXiv:1310.6279 [math.PR] to appear in J. Appl. Probab., Dec. (2014).
25. Li, S.: Concise formulas for the area and volume of a hyperspherical cap. Asian J. Math. Stat. 4(1), 66–70, (2011)
26. Lord, R.: The Distribution of Distance in a Hypersphere. Ann. Math. Statistics 25(4), 794–798, (1954)
27. Mathai, A. M.: A handbook of generalized special functions for statistical and physical sciences. Oxford Science Publications, Oxford, (1993)
28. Olver, F.W.J., Lozier, D.W., Boisvert, R.F., Clark, C.W. editors. NIST handbook of mathematical functions. Cambridge University Press, Cambridge, (2010)
29. Orsingher, E., de Gregorio, A.: Random Flights in Higher Spaces. J. Theoret. Probab. 20(4), 769–806, (2007)
30. Pearson, K.: The problem of the random walk. Nature 72, 294, (1905); 72, 342, (1905)
31. Pearson, K.: A mathematical theory of random migration. Mathematical Contributions to the Theory of Evolution XV. Draper's Company Research Memoirs, Biometric Series. Dulau and co., London, (1906)
32. Pogorui, A., Rodríguez-Dagnino, R.M.: Random Motion with Uniformly Distributed Directions and Random Velocity. J. Stat. Phys. 147(6), 12161225, (2012)
33. Pogorui, A., Rodríguez-Dagnino, R.M.: Random motion with gamma steps in higher dimensions. Statist. Probab. Lett., 83(7), 1638–1643, (2013)
34. Stadje, W.: The exact probability distribution of a two-dimensional random walk. J. Stat. Phys. 46(1-2), 207–216, (1987)
35. Stannard, A., Coles, P.: Random–walk statistic and the spherical harmonic representation of CMB Maps. Mon. Not. R. Astron. Soc. 364, 929-933 (2005)
36. Viswanathan, G. M., da Luz, M. G. E., Raposo, E. P., Stanley, H. E.: The Physics of foraging. An introduction to random searches and biological encounters. Cambridge University Press, Cambridge, (2011)
37. Watson, G.N.: A treatise on the theory of Bessel functions. Cambridge Mathematical Library. Cambridge University Press, Cambridge, (1995)

38. Zoia, A., Dumonteil, E., Mazzolo, A.: Collision densities and mean residence times for d-dimensional exponential flights. Phys. Rev. E, 83(4), 041137, (2011)